

Variability of Ground Motions in Southern California—Data from the 1995 to 1996 Ridgecrest Sequence

by Craig W. Scrivner* and Donald V. Helmberger

Abstract Seismograms from the 1995 to 1996 Ridgecrest, California, earthquake sequence, recorded by the TriNet digital seismic network, provide high-quality waveforms from sites throughout southern California, including sites in markedly heterogeneous areas like the Los Angeles area sedimentary basins. Synthetic seismograms calculated by the reflectivity method with various 1D models are used as a baseline to measure the variability of amplitudes throughout southern California. Regardless of the model used, there is greater variability in the amplitudes from basin site records than from rock site records. Rock, soil, and basin sites are all rather insensitive to radiation pattern nodes at the three frequency bands investigated: 0.1 to 0.2 Hz, 0.2 to 0.4 Hz, and 0.4 to 0.8 Hz. This complicates the analysis because the nodes create singularity points in the distribution of ratios of observed and synthetic amplitudes. When stations near nodal planes are removed, the surface waves observed at most rock sites have peak amplitudes within a factor of 2 of synthetic waveform amplitudes. Peak amplitude of the surface waves observed at the soil and basin stations are more variable, with the bulk of the distribution of data/synthetic amplitude ratios less than 3 and a few outliers greater than 5. These outliers occur at the higher frequency bands. Soil and basin sites are also more often larger than the synthetics (higher median values). Most outliers can be explained by applying a water level of 50% to the radiation pattern. This reduces the scatter in the distributions to about the same extent as removing data within 10° of nodes. Thus, most of the outliers are sites that are insensitive to the nodes, not sites that are larger than the overall data distribution.

Introduction

Earthquakes north of Ridgecrest, California, were the first sequence of moderate events to be recorded by the strong-motion component of the TriNet digital network. This network is unique in that basin stations are included to gain information about the basin response in addition to the usual hard-rock siting normally used in earthquake source studies. The network is not complete, but many stations were installed during the Ridgecrest sequence, mid-August 1995 to early-January 1996 (Table 1). The data set is unusual in that it includes records of surface waves from a moderate earthquake as they pass through the Los Angeles basins (Los Angeles, San Fernando, San Gabriel, and San Bernardino). The events are far enough from the basins that surface waves have already developed before reaching the basins. This allows a comparison of surface waves on records at stations inside and outside the basins, which is the subject of this article (Fig. 1).

In this data set, we find incident surface waves generating secondary surface waves in the basin. For example, in Figure 2a, tangential components of records from the 17 August 1995 event at the rock site PAS and the nearby basin station KIK are shown to be quite similar for the first 10 to 15 sec after the initial shear-wave motion. At the time where a large second phase arrives at PAS, the same phase arrives at KIK, and coda begins that is not seen at PAS. Similar basin-generated waves are seen in the comparison of records ob-

Table 1
Events Analyzed in this Study
Time and location were determined by the Southern California Seismic Network.

Date	Time	Lat	Lon
17 August 1995	22:39:59	35.78	–117.66
20 September 1995	23:27:36	35.76	–117.64
7 January 1996	14:32:54	35.77	–117.65

*Present address: California Division of Mines and Geology, 801 K Street, MS 13-35, Sacramento, California 95814-3531.

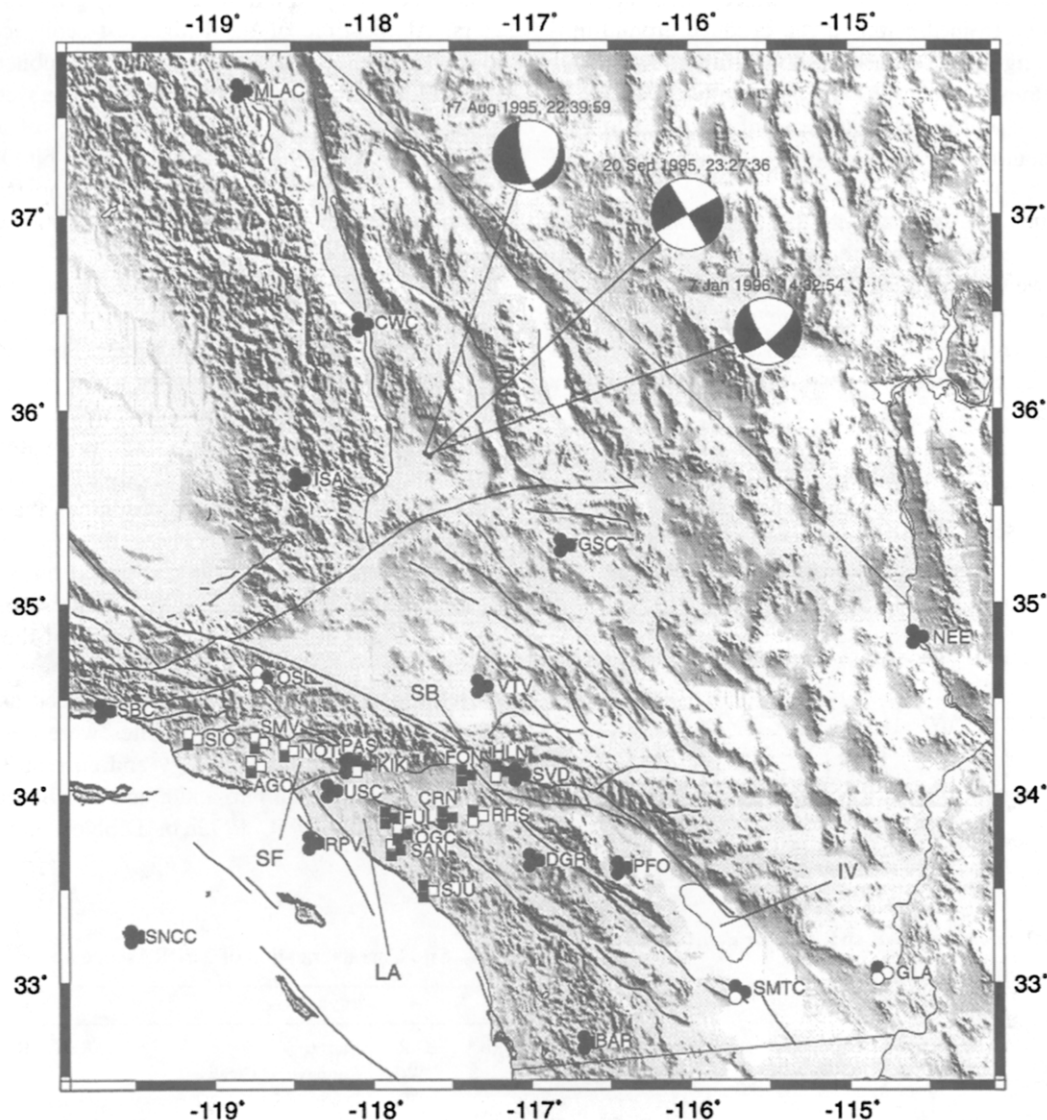


Figure 1. Events examined in this study and the stations at which records are available. Circles indicate TERRAScope very broadband velocity receivers. Squares indicate K2 force balance accelerometers. Three symbols are shown for each station, one for each event. A black (white) symbol indicates the event was (was not) recorded by that station. The thick black line is the California coast and border. The thin black lines are some significant faults in southern California. The gray lines are highways. Also labeled are the San Fernando (SF), Los Angeles (LA), San Bernardino (SB), and Imperial Valley (IV) sedimentary basins mentioned in the text.

served at the soil station CRN and basin stations FUL, OGC, and SAN in Figure 2b. In these examples, the amplifications at the basin stations are greatest at higher frequencies.

Surface waves in basins are of engineering interest because they are often the largest arrivals at a site, they are of long duration, and the frequency content of the shaking is similar to the resonance frequencies of tall buildings, bridges, and other large, flexible structures. In southern California, the San Andreas fault has the potential to generate the largest earthquakes and is far enough away from the Los Angeles area that surface-wave amplification by the basins

is an important concern. Olsen *et al.* (1995) have simulated an $M 7.75$ earthquake on the San Bernardino segment of the San Andreas fault and the subsequent propagation of energy into the Los Angeles and San Fernando basins. Their simulation produces much larger spectral amplitudes in the basins than at surrounding sites (up to 10 times higher at 4- to 5-sec period). Teng and Qu (1996) simulated long-period ground motions and strain distribution from an $M 8.25$ earthquake on the Mojave segment of the San Andreas fault. They find peak amplitudes up to 3 m at 3- to 10-sec period and localized strains of 10^{-2} to 10^{-3} .

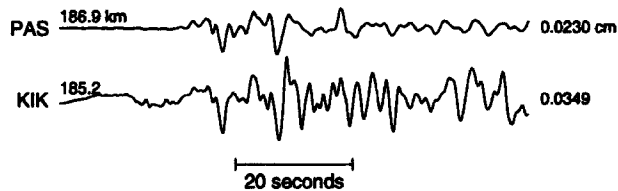
These simulations suggest that surface waves at 60 to 100 km distance from the source can produce ground motions large enough to be of great engineering concern. Our confidence in these simulations and their applicability to engineering design will increase as they are grounded in data and an understanding of the variability in recorded ground motions.

In this study, we will analyze these new data in terms of our experience gained from hard-rock modeling studies. In particular, we will examine the frequency-dependent amplitudes of the various stations as compared against 1D synthetics so that the first-order effects of radiation pattern and propagational effects can be assessed independently of the basin amplification.

Data

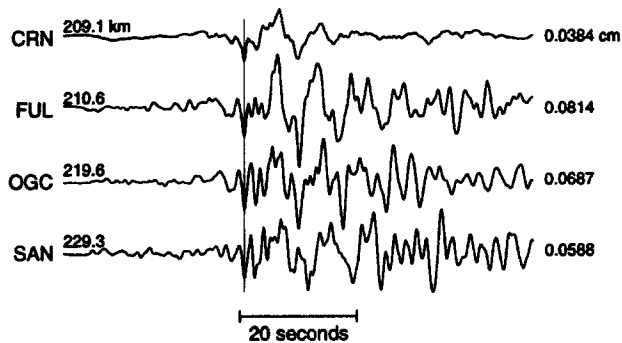
The TriNet seismic array consists of three subnets with a small number of very broadband receivers, many broad-

95/08/17 Tangential



(a) Tangential records at PAS and KIK for event 95/08/17

95/09/20 Tangential



(b) Tangential records at CRN, FUL, OGC, and SAN for event 95/09/20

Figure 2. Tangential component records from stations PAS and KIK for the 17 August 1995 event and stations CRN, FUL, OGC, and SAN for the 20 September 1995 event. For each event, the waveforms are plotted on the same amplitude scale. The PAS and KIK waveforms are plotted with absolute timing. The CRN, FUL, OGC, and SAN waveforms are aligned on the phase indicated by the vertical line. The range and peak amplitude are indicated at the left and right ends of each trace, respectively.

band receivers, and a number of strong-motion accelerometers. At the time of the Ridgecrest sequence, data were available from the TERRAscope and K2 subnets (Fig. 1).

TERRAscope is the very broadband subset of the complete array. There are 20 sites consisting of a Streckeisen STS-1 or STS-2 velocity sensor and a 24-bit Quanterra data logger. The data are sampled at 20 samples/sec. The response is flat in velocity from 100 sec (300 sec for STS-1 sensors) to 7 Hz.

K2 is the strong-motion subset of the complete array. The instrument package consists of a Kinematics three-component FBA and 19-bit data logger. The data are sampled at 100 samples/sec. The response is flat in acceleration up to 25 Hz. At the time of the Ridgecrest events, the strong-motion subnet was still under installation, with about 10 stations, and the handling of the data was not completely automated. As a result, the particular stations available for each event changes, and in some cases, the available waveform ends before the coda is complete. Some records from stations CRN and KIK have large long-period noise in the *P*-wave portions of the waveform. This portion of the records was clipped before finding peak amplitudes.

Site conditions for stations are defined as either rock, soil, or basin (Table 2). Basin sites were designated by a combination of surface geology and current 3D models of the Los Angeles, San Fernando, and San Bernardino basins (Dibblee 1968, 1989, 1992a,b; Dibblee and Ehrenspeck,

Table 2
Site Characterization of TERRAscope and K2 Stations

	TERRAscope		K2	
Rock	BAR	Barrett	RRS	Riverside
	CWC	Cottonwood Creek		
	DGR	Domenigoni Valley		
	GLA	Glamis		
	GSC	Goldstone		
	ISA	Lake Isabella		
	NEE	Needles		
	OSI	Osito Canyon		
	PAS	Pasadena		
	PFO	Pinon Flat		
	RPV	Rancho Palos Verdes		
	SMTC	Superstition Mountain		
	SNCC	San Nikolas Island		
	VTV	Victorville		
Soil	MLAC	Mammoth Lakes	AGO	Agoura
	SVD	Seven Oaks Dam	CRN	Corona
			SJU	San Juan Capistrano
			SMV	Simi Valley
Basin	SBC	Santa Barbara	FON	Fontana
	USC	University Southern California	FUL	Fullerton
			HLN	Highland
			KIK	Kinematics
			NOT	Northridge
			OGC	Orange
			SAN	Santa Ana
			SIO	Ventura County

1993; Greenwood and Morton, 1991; Irvine, 1990; Magistrale *et al.*, 1996; Morton, 1978a,b; Morton and Cox, 1994). The strong-ground-motion attenuation relationships usually do not include a separate classification for basin sites (e.g., Boore *et al.*, 1997). Normally soil sites are distinguished by the characteristics of the top few tens of meters beneath the surface. Here we are concerned with the difference between deep (greater than about 1 km) basins and the surrounding crust.

The TERRAscope data was highpass filtered with a corner at 0.05 Hz and integrated to displacement. The K2 data were handled in the same way, except that an additional step of integration was required going from acceleration to displacement. For the 17 August and 20 September 1995 events, a K2 station was maintained at the same site as the Pasadena TERRAscope site. Figure 3 compares the traces for the two instrument packages after the processing described earlier. The K2 traces are shorter than their TERRAscope counterparts, as previously mentioned, and contain long-period noise. Otherwise, the wave shapes are identical. The waveforms are all plotted on the same scale. The K2 records have somewhat lower peak amplitudes than the TERRAscope records (except the tangential component of event 20 September 1995). This variation is within 10% of the amplitude of the TERRAscope tangential and radial traces and within 25% of the amplitude of the TERRAscope vertical traces.

Analysis

We use hard-rock sites from the TERRAscope subarray of TriNet as the background wave field against which to compare basin stations. The comparison is indirect, using seismograms calculated by the reflectivity method using 1D velocity models as a reference. This allows comparison of data from stations at a variety of distances and azimuths from the sources.

The velocity models used in the analysis are a standard model of southern California (Dreger and Helmberger, 1991), and a variant with a slow surface layer (Table 3). The standard model was used in inversions of TERRAscope data by Zhu and Helmberger (1996) for earthquake source parameters throughout southern California. In our inversions for source parameters and our comparisons of data from hard rock, soil, and basin sites, we apply both the standard model and the variant model with a slow surface layer. Synthetic waveforms are calculated for these models by the reflectivity method (Saikia, 1994). The top layer in the variant model has $Q_p = 100$ and $Q_s = 50$. Synthetic waveforms were also calculated with this layer set to $Q_p = 600$ and $Q_s = 300$. The change in peak amplitude with attenuation was negligible.

The inversion technique to determine source characteristics of the three earthquakes is a grid search over the parameter space (Zhao and Helmberger, 1994). The extended P waves (P_{nl}) and surface-wave segments of the waveforms are handled separately in the inversion. Each segment is al-

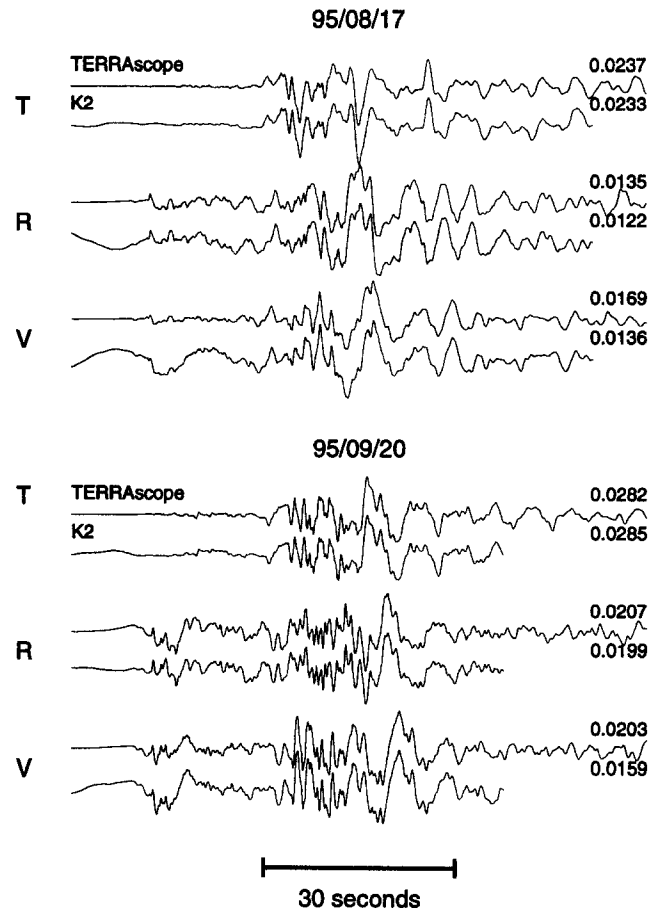


Figure 3. TERRAscope and K2 waveforms from the Pasadena site for the 17 August 1995 and 20 September 1995 events. The waveforms are all plotted on the same amplitude scale. The peak amplitude is indicated at the right end of each trace.

Table 3

One-Dimensional Models

Z is the depth to the top of the layer in kilometers. V_p and V_s are the compressional and shear-wave velocities in km/sec. ρ is the density in g/cm³. Q_p and Q_s are compressional and shear-wave attenuation factors.

Z	Standard					Variant				
	V_p	V_s	ρ	Q_p	Q_s	V_p	V_s	ρ	Q_p	Q_s
0.0	5.50	3.18	2.40	600	300	4.00	2.31	1.80	100	50
3.0						5.50	3.18	2.40	600	300
5.5	6.30	3.64	2.67	600	300	6.30	3.64	2.67	600	300
16.0	6.70	3.87	2.80	600	300	6.70	3.87	2.80	600	300
35.0	7.80	4.50	3.00	600	300	7.80	4.50	3.00	600	300

lowed small, independent shifts in timing. The waveforms are lowpass filtered with a corner period of 2 sec. Absolute amplitudes are kept in the inversion, with a weighting scheme to evenly emphasize data from different distances (Zhu and Helmberger, 1996). The inversion with absolute amplitudes is more stable near nodes in the data. A 2-sec

triangular source time function is convolved with the synthetic waveforms. An application of this technique to the largest event in the Ridgecrest sequence (20 September 1995) is presented in Zhu and Helmberger (1996). A comparison of the surface-wave portions of the data against synthetics for the first event (17 August 1995) is given in Figure 4.

Examining all the data from one event, as in Figure 4, there is a strong contrast between the duration of shaking at rock sites and basin sites. Predominant period of seismograms at most rock sites is relatively short (10 to 20 sec). Stations with nodes of Love waves (e.g., PAS and RPV) are emergent, with the peak amplitude 10 to 15 sec behind the direct shear wave. The radial component is more variable in amplitude and waveform than the tangential component. In most cases, the Rayleigh waves dominate the records, but at some stations (for example, VTV), high-frequency S_{nl} waves are also strong.

Basin sites have longer durations than rock sites. The long-period surface waves extend 40 sec or more. Coda on the radial component tends to have higher frequency content than coda on the tangential component. Coda amplitudes on both components are often as large as the main surface waves. Records from soil sites vary greatly. They tend to mimic one of the two extremes, rock or basin.

Particularly notable waveforms are those from SMTC, SJU, FON, and HLN. Both SMTC and SJU show signs of the wave field's interaction with neighboring basins. SMTC is just southwest of the Imperial Valley. The long duration of large, long-period surface waves at SMTC is in sharp contrast to the records at surrounding sites PFO, BAR, and GLA, all of which skirt the Imperial Valley. Ho-Liu and Helmberger (1989) and Helmberger *et al.* (1992) modeled waveforms from earthquakes in and around the Imperial Valley as recorded at Pasadena. They found that surface waves generated in the Imperial Valley passed into the surrounding crust and persisted to large distances as strong coda. Our data indicate that the reciprocal path is just as effective in trapping surface waves. The entire path consists of surface waves striking the basin, the generation of additional surface waves in the basin, and the entire package of surface waves propagating from the basin to the down-range site of SMTC. SJU is south of the Los Angeles basin on Tertiary marine sedimentary rock. The period of the coda at this station is similar to that seen in southern Los Angeles basin stations FUL and SAN. The coda also seems to have similar duration and amplitude of coda as the basin stations, but unfortunately, the records at SJU are cut off before the coda ends. These two sites indicate that surface waves generated in basins can propagate, with little apparent diminution, to sites beyond the trailing edge of the basin.

FON and HLN seismograms show relatively little basin effect. The sedimentary basement interface of the San Bernardino basin and the alluvium deposits around FON are shallower than the San Fernando and Los Angeles basins and may not be deep enough to cause effective trapping of

the 5- to 10-sec surface waves that dominate records from the rock and soil sites around it (VTV and SVD).

Note that in Figure 4, the Love wave node falls near stations RPV and PAS, while DGR should have large-amplitude Love waves. Instead, the data amplitudes for all three stations are very similar. In this case, high frequencies are not strong at any of the stations. In contrast, NEE is expected to be affected by a node, but the peak amplitude of surface waves on the tangential component is as large as that at PFO. However, NEE is particularly noisy at high frequencies.

Source Parameters

For each event, three inversions were done for source-mechanism parameters strike, rake, dip, moment, and depth (Table 4). In the first inversion, only seismograms observed at the rock TERRAScope sites were used to examine the background wave field. The parameters found for the first two events are very similar to the inversions done by Zhu and Helmberger (1996). The second inversion was based on all the sites in our analysis. Before running this inversion, a series of single-station inversions were done with each solution parameter (strike, rake, and dip) constrained to within 10° of the parameter value from the first inversion. Each single-station inversion was run for both the standard and variant models, and the model with the smaller error was selected as the preferred model for that station. Then the full inversion was run based on all the data, with each station fitting its preferred model. The synthetics for this solution for the 17 August 1995 event are shown in Figure 4 as an example. For each event, this second source solution was very similar to the first inversion solution. Most of the data added in this inversion are from stations in the Los Angeles area, clustered around a Love-wave node in the original inversion solution. The similarity of the second inversion to the original inversion increases our confidence that the inversion is accurate, including the location of nodes in the radiation pattern.

In these inversions, the moment is controlled by the peak amplitudes of the waveform segments. Thus, surface-wave amplitudes will be fit by synthetics with the moment from the inversion. Because of variations in amplitude from site to site, however, there is scatter around this average fit. An error in the source depth estimate would change the moment obtained by the inversion and affect the relative amplitude of the P_{nl} and surface-wave segments of the synthetic waveforms. Our results are insensitive to these effects because we are focusing on surface waves and looking at ratios of data to synthetic. To check the depth estimate of the full inversion, a third inversion was run using only the P_{nl} portion of the waveforms. This inversion returned very similar focal mechanism parameters, including depth and moment, to the previous inversions for all three events.

Amplitude Variation

In this section, we compare peak amplitudes of the surface waves observed at the hard-rock, soil, and basin sites

China Lake 95/08/17 22:39:59
Displacement
Rock Sites

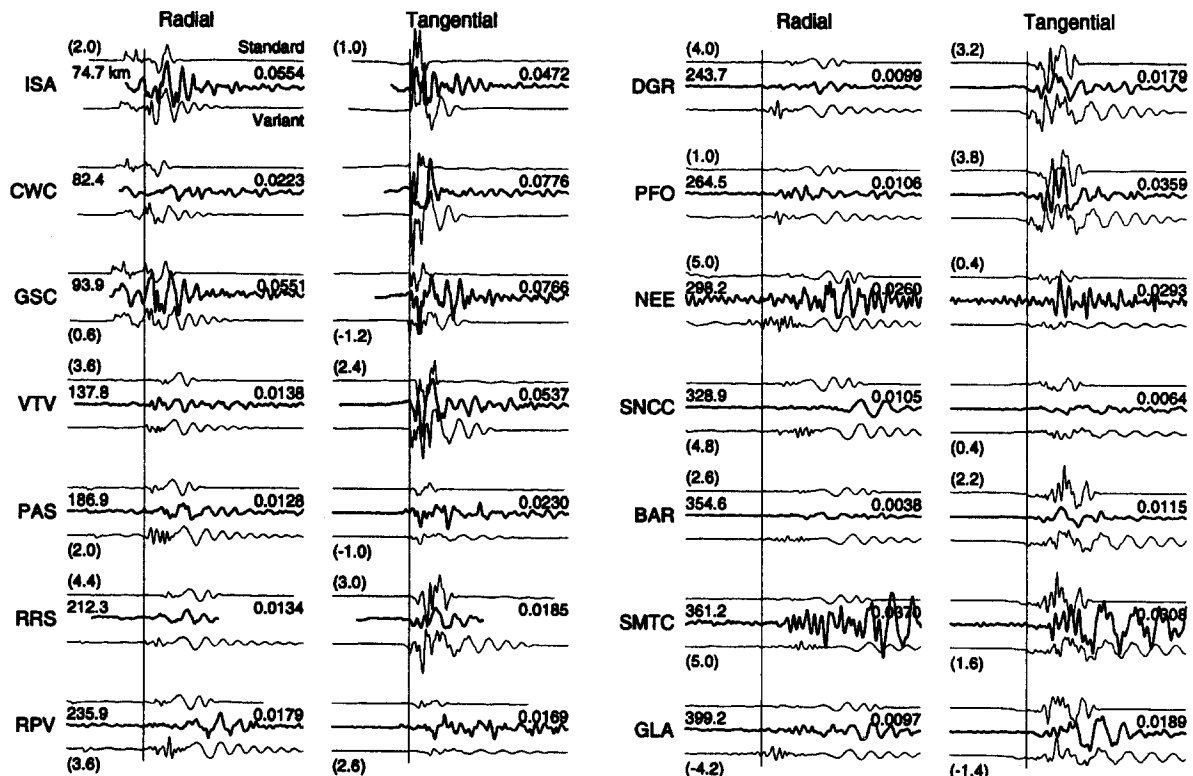


Figure 4. Displacement records of the 17 August 1995 Ridgecrest earthquake, organized by site characteristics. Source parameters for synthetics are from an inversion of the complete data set for the event. For each station, the three waveforms are standard model synthetic (top), data (middle), and variant model synthetic (bottom). The data are aligned by the time of the first theoretical shear-wave arrival for the standard model (vertical lines). The synthetic waveforms are plotted with a time shift that maximizes the cross-correlation between observed and synthetic seismograms. The shift is listed in parentheses at the left end of the synthetic in units of seconds, positive shift to the right. The amplitudes of the waveforms are scaled by a factor of $(r/100)^{0.5}$, where r is the range in kilometers. The peak amplitude (cm) of the data is shown at the right end of the trace.

Table 4
Events analyzed in this study with source inversion results with different subsets of the data.
The inversion technique is due to Zhao and Helmberger (1994).

Event	Dataset	Strike	Rake	Dip	Moment	Depth
17 August 1995	TERRAscope sites, full waveform	50	330	30	1.6×10^{24}	11
	all sites, full waveform	54	328	50	7.9×10^{23}	11
	all sites, <i>Pnl</i> segments only	50	324	50	1.1×10^{24}	11
20 September 1995	TERRAscope sites, full waveform	60	360	90	2.2×10^{24}	14
	all sites, full waveform	60	340	66	1.6×10^{24}	14
	all sites, <i>Pnl</i> segments only	62	334	66	1.6×10^{24}	14
7 January 1996	TERRAscope sites, full waveform	50	340	70	5.6×10^{23}	8
	all sites, full waveform	58	338	74	2.8×10^{23}	8
	all sites, <i>Pnl</i> segments only	58	342	64	2.8×10^{23}	8

in three pass bands (0.1 to 0.2 Hz, 0.2 to 0.4 Hz, and 0.4 to 0.8 Hz) and examine their variability. We examine different pass bands because our initial analysis of the data, such as the examples we discuss in the Introduction, suggested frequency dependence of the amplification. The low-frequency pass band is controlled by long-period noise in the data. The high-frequency pass band is controlled by the limitation of using synthetic waveforms calculated with 1D models.

In Figure 5, the peak amplitudes of surface waves observed at epicentral distances of 185 to 253 km from the 20 September 1995 event are compared to the amplitudes of the synthetic waveforms at an epicentral distance of 220 km for the standard model. The data are classified by site type and frequency bandpass. The range of the data examined here is limited to the region around the Los Angeles basin region. The 20 September 1995 event has a pure strike-slip mechanism, with a tangential node among the stations in this distance range, so the contrast between the synthetic and data amplitudes at the node can be clearly examined. The variation in peak amplitude expected across the distance range of 185 to 253 km from the synthetic waveforms is about 0.01 cm, which is much less than the scatter in the data amplitudes on the horizontal components, thus the comparison of the data with synthetic waveforms calculated at one range is reasonable. Later we use synthetics for each specific site to compare amplitudes over the whole data set.

On the horizontal components, the peak amplitudes of the basin stations, in all three frequency bands, are larger than those of the rock and soil sites. On the radial component, the high-frequency pass band has particularly large peak amplitudes. The vertical component has less variation between types of sites and across the frequency pass bands.

In general, the synthetic amplitudes from the standard model (Fig. 5) are comparable to the data amplitudes, except on the radial component. On the tangential component, data from the rock and soil sites east of the node (160° to 180°) are smaller than the synthetic predicts. Around the node, data from the basin sites are larger than predicted by the synthetics, though the amplifications are slightly depressed at the node. There is no clear frequency dependency for this

insensitivity of the data to the node. On the radial component, the rock- and soil site amplitudes are fit well, but the basin sites are 2 to 4 times higher (and even more at higher frequencies). Only a few stations are near nodes on the radial and vertical components, but from these data, it appears that data on the vertical component follow the nodes more closely than data on the radial component.

These are qualitative comparisons of a subset of the data from one event. In order to include the entire data set, with ranges from 75 km (station ISA) to 400 km (station GLA) and a wider range of azimuths, we take the ratio of the peak amplitude of the data and the synthetic waveform for each station. Because the seismic moment in the synthetics is based on the inversion, the average ratio of the peak amplitude of the data and the synthetics should be about 1. However, the inversion is done with the long-period end of the data (greater than 5 sec), so there will be shifts away from 1 in the higher frequency pass bands used in the ratios.

To summarize the comparison of data and synthetic amplitude ratios for all three events as a single data set, the distributions of ratios are presented as box plots (Devore, 1987). Each component and pass band is handled as a separate data set. The box-plot format (Fig. 6) was chosen because it presents the entire data set while indicating statistical features and the shape of the distribution. For the ratios of data/synthetic amplitude, the ratios tend to be skew, and using mean and standard deviation as statistical measure obscures this feature. Also, the outliers are an important portion of the data for engineering applications. We label outliers and discuss them later in the article.

Figure 7 shows the data for the three events at all sites combined into one data set. When the standard model is used to calculate synthetic waveforms (7a), the bulk of the distribution is down around 1, except the higher frequencies on the radial component. The scatter is large on the tangential and radial components. Use of the variant model synthetics in the ratios (7b) brings the median of the distributions on the radial component down around 1, but the scatter is not reduced much, as is also the case for the mixed path model. It might seem that the variant model, with a 3-km-thick top

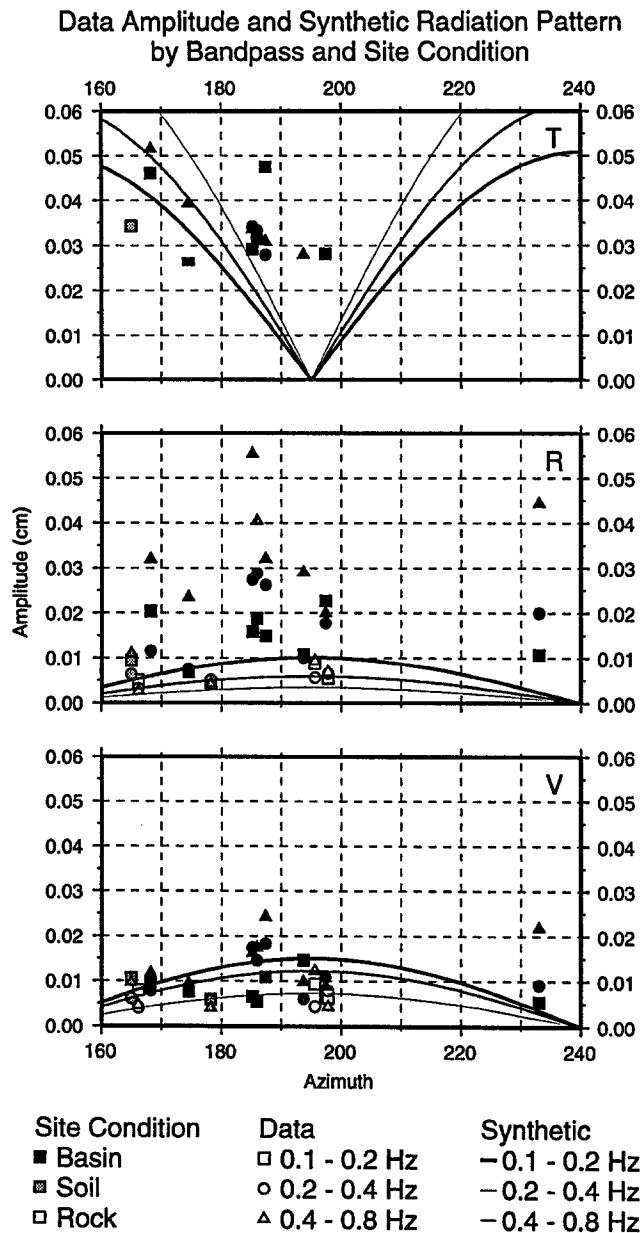


Figure 5. Maximum amplitudes of a subset of data from the 20 September 1995 event and synthetic waveforms from the standard model. Station distances range from 185 to 253 km. The synthetic waveforms were calculated for 220 km distance. Data are discrete points; synthetics are curves. Amplitudes are shown for three different bandpasses of the waveforms, with the square/circle/triangle symbol indicating a bandpass of (0.1 to 0.2)/(0.2 to 0.4)/(0.4 to 0.8) Hz, respectively. Shading of black, gray, and white indicates a basin, soil, and rock site condition, respectively. The top plot is the tangential component, the center plot is the radial component, and the bottom plot is the vertical component.

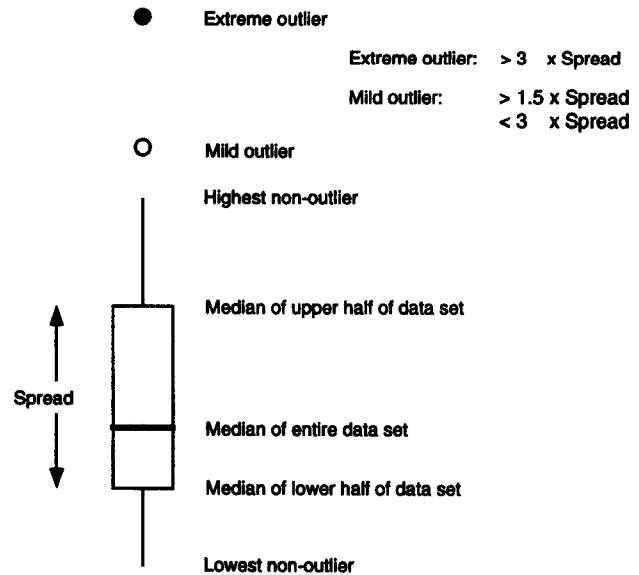


Figure 6. Box-plot construction and parameters. The thick horizontal line in the box is the median of the data. The top and bottom of the box are the medians of the upper and lower halves of the data, respectively. The length of the box is the *spread* of the data, and it is used to define outliers in the data. A *mild* outlier is farther than $1.5 \times \text{spread}$ from the top or bottom of the box. An *extreme* outlier is farther than $3 \times \text{spread}$ from the top or bottom of the box. The vertical lines protruding from the box ends are the highest and lowest nonoutlier values. Mild outliers are indicated by open circles; extreme outliers, by filled circles.

slow layer, should only be applied to the basin data. With this approach (7c), the radial component ratio distributions are wider, with more outliers, than when the variant model is used for all the sites. This suggests that the variant model is preferable for all three site conditions.

It is evident that data/synthetic amplitude ratios near nodes in the surface-wave radiation pattern are unstable because amplitudes in synthetic waveforms may fall on a node. The theoretical minima are less extreme when the radiation pattern has an oblique component. Thus, the correspondence of data with theoretical nodes depends on path complexities and multipathing. When the sites within $\pm 10^\circ$ from surface-wave nodes are removed from the amplitude ratio data set (Fig. 8a), the number of outliers is significantly reduced. This is clearest on the tangential component, which has a node in the Los Angeles basin area for all three events. On this component, the spread of the distribution is also reduced. This implies that, away from nodes, the bulk of the data at all site conditions can be explained to within a factor of 2 by a 1D model. However, there are individual sites where the data amplitudes are more than three times greater than predicted by the synthetic waveforms.

If the majority of the outliers in the complete data set (Fig. 7) can be explained by the insensitivity of the data to

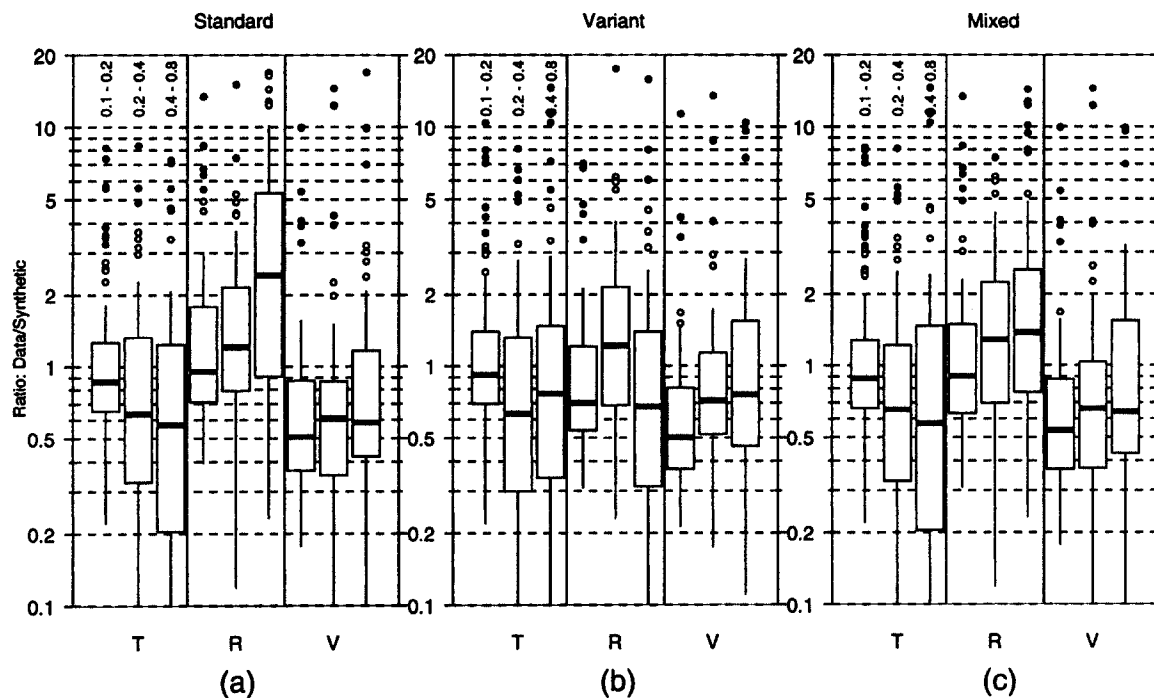


Figure 7. Distribution of data/synthetic ratios for all sites that recorded the 17 August 1995, 20 September 1995, and 7 January 1996 events. Separate statistics are shown for the 0.1 to 0.2, 0.2 to 0.4, and 0.4 to 0.8 Hz frequency bands. (a) Synthetic waveforms are generated with the standard model, (b) synthetic waveforms are generated with the variant model, and (c) a mixed model with the standard model used to generate synthetics for the rock and soil sites and the variant model used to generate synthetics for the basin sites.

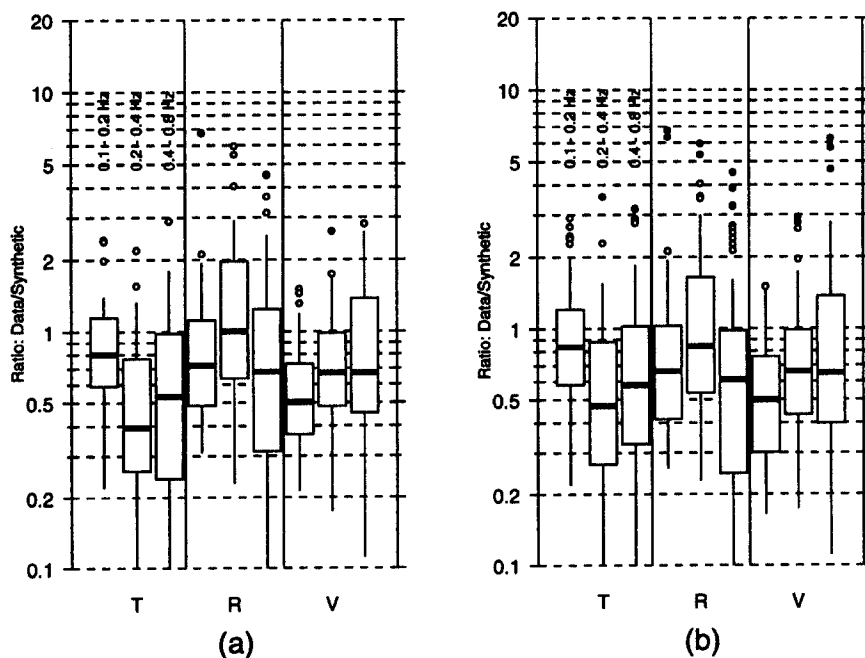


Figure 8. Same as Figure 7b except that in (a) data within $\pm 10^\circ$ of source radiation nodal planes are removed and in (b) a water level of 50% is applied to the radiation pattern of the synthetic waveforms. In both cases, the synthetic waveforms are calculated with the variant model.

source nodes because of weak multipathing effects, then adding a water level to the source radiation pattern should have a similar effect on the distribution of ratios as removing sites near nodes. Adding a water level simply means forcing the synthetic amplitude used in the data/synthetic ratio to be at or above a percentage of the overall maximum amplitude of the synthetic waveform at all azimuths. A water level of 50% (Fig. 8b) has an effect similar to removing data within 10° of the nodes (Fig. 8a). Outliers exist on all three components and at almost all of the frequency bands examined, however, and this suggests that there are sites where moderate amounts of multipathing cannot explain the data amplitude.

Next, we re-examine the data/synthetic ratios for each site condition independently to isolate amplifications at non-rock sites. Ratios from rock sites are shown in Figure 9a. The distribution of the data/synthetic amplitude ratios is fairly compact. On the tangential and vertical components, the bulk of the ratios is at or below 1, with a few extreme outliers. There is more scatter on the radial component, particularly in the 0.4- to 0.8-Hz bandpass. Using the variant model in the ratios has the greatest effect on the radial component, reducing both the number of outliers and the spread of the distribution as discussed earlier. Applying a water level of 50% to the radiation pattern reduces the number of outliers even more (lower panel), in particular extreme outliers. The largest ratio of data to synthetic amplitude becomes less than 3, except for one extreme, SMTC, as labeled. This particular station is probably contaminated by energy trapped in the Imperial Valley. We consider this station further in the Discussion section.

Compared to the rock sites, the ratios of data/synthetic at the soil sites (Fig. 9b) have fewer outliers. The ratios for the radial component have wider distributions, particularly the 0.4- to 0.8-Hz bandpass, with higher median values. Use of the variant model synthetics in the ratios considerably narrows the distribution for the 0.4- to 0.8-Hz bandpass and reduces the median. It has little effect otherwise. It is worth noting that there is less data from soil sites than the rock and basin sites. There are 38 records from rock sites, 13 records from soil sites, and 22 records from basin sites. For the soil sites, this has the effect of reducing the number of outliers in the distribution, as defined for the box plot, while having wide distributions. Removing data around nodes or applying a water level to the source radiation pattern (lower panel) has little effect on the distributions; however, the number of extreme outliers is reduced.

At basin sites (Fig. 9c), the distributions for all three bandpasses are relatively wide. Some of this is due to the location of the Love-wave nodes for the three events. They fall on about the same azimuth, through the Los Angeles basin, and a relatively high percentage of the basin sites are near these azimuths. As with the rock and soil sites, the ratio distributions for the radial component are wider than for the vertical component, and higher frequency distributions are wider. Use of the variant model for the synthetics has a simi-

lar effect on the basin ratios as on the rock and soil sites, reducing the scatter on the radial component but having little effect on the tangential and vertical components. Applying a water level to the source radiation pattern (lower panel) has a significant effect on the results for the tangential component. This reflects the locations of the Love-wave nodes. Even with the water level applied, there are still large outliers at SAN, FUL, SBC, and SIO.

A comparison of the rock and basin sites indicates that the rock sites tend to have lower amplitudes and show less variability than basin sites. With data around nodes removed, rock sites tend to have data/synthetic amplitude ratios at or below 1, in all three frequency bands. The largest outliers (except SMTC) are less than 3. In fact, the synthetic amplitudes are often overestimation, and ratios approach 0. Ratios at basin sites tend to fall between 1 and 2. With data around nodes removed, only a few outliers are larger than 3. However, the extreme outliers from SIO are as high as a factor of 9.

Discussion

When examining amplifications, outliers are quite significant, because they suggest that specific sites are so anomalous that they require special engineering concern or that a type of site condition or location is likely to receive stronger ground motions. A well-known example is the Tarzana site that recorded high accelerations in the 1994 Northridge mainshock (1.8 g). Other stations within 2 km of the site recorded significantly smaller accelerations. Two recent efforts to explain the amplification at Tarzana have argued for resonance of the (fairly subdued) topography of the hill where the instrument is set (Spudich *et al.*, 1996) and energy from a small landslide directly under the station, triggered by the mainshock shaking (Rial, 1996). Rial's model is the more exotic of the two, but Spudich *et al.*'s model creates more potential problems for seismic hazard mapping because it implies that subtle and common focusing effects may have to be taken into account.

In our data set, after nodal points are removed or a 50% water level is applied to the radiation pattern, ratios are greater than 3 at only a few specific stations: SMTC (rock); SJU and MLAC (soil); and SAN, FUL, SBC, and SIO (basin). MLAC is near a Rayleigh-wave node, but when the water level is applied, it is still anomalous. This site is in the unusual geologic environment of Long Valley Caldera that is likely to produce strong scattering of incoming seismic energy and to trap seismic energy in slow near-surface materials. SAN, FUL, and SBC are at the surface of large sedimentary basins. SAN and FUL are at the southern end of the Los Angeles basin. SBC is in the Santa Barbara basin. Earlier it was noted that seismograms at those three sites show large basin-generated surface waves, by comparison to a nearby soil station (Fig. 2). However, the anomalous amplitudes of the records, generating the outliers, are at high frequencies, rather than due to the longest periods in the surface waves.

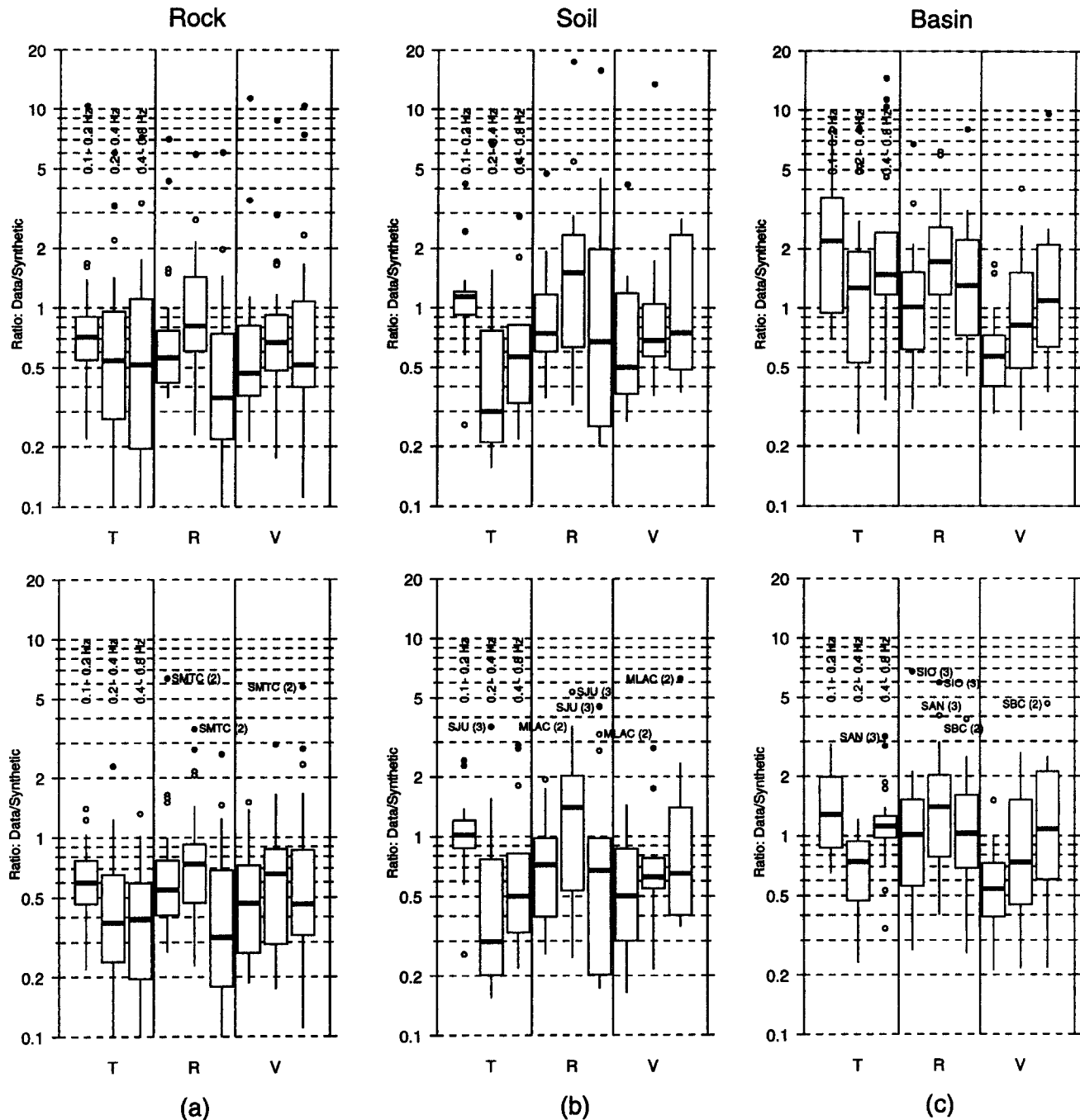


Figure 9. Distribution of data/synthetic ratios for each site condition. The synthetic waveforms are calculated with the variant model. Direct ratios of data and synthetic amplitudes (equivalent to Fig. 7b) are shown in the upper panels. Ratios for which the synthetic amplitude is adjusted with a 50% water level (equivalent to Fig. 8b) are plotted in the lower panels: (a) rock sites, (b) soil sites, and (c) basin sites.

As expected from the large, long-period coda at SMTC and SJU, amplitude ratios at these stations are very large. These two stations are located just beyond the Imperial Valley and Los Angeles sedimentary basins, respectively. This suggests the large amplitudes and long codas can be explained as basin-generated surface waves leaking from the trailing edges of the basins. Bard and Bouchon (1980) noted

that basins with low-velocity contrast to the background structure are ineffective at reflecting surface waves back into the basin. This appears consistent with the comparison of waveforms from SJU and stations in the southern end of the Los Angeles basin (SAN, OGC, and FUL). The amplitude and duration of the coda at SJU is very similar to that recorded at the basin stations.

The SIO seismograms are the most anomalous in this data set. Only the radial component has large amplifications, and the amplification is extremely large and occurs at long periods. We have not resolved why this station is so anomalous. The station is at the edge of the deep Ventura basin, and edge effects might explain the large amplitudes. However, the extreme contrast between components suggests instrumentation problems. Unfortunately, SIO only recorded the third event, and the ground motion cannot be compared for different events.

Stations tend to have anomalously large amplitude ratios for one event but not the others. The only exception to this is station SJU, which has outlier points for both events 17 August 1995 and 7 January 1996. The correlation of outlier stations with specific events suggests a source effect in the generation of the large amplitudes. One possibility is that the amplitudes of basin-edge surface waves are relatively sensitive to incident surface waves that are themselves sensitive to source parameters, particularly near nodes. In this regard, it is interesting that while FUL and SAN have outliers for different events, SJU has outliers for both those events. Since SJU is receiving leaked basin surface waves generated throughout the Los Angeles basin, the site may be less sensitive to nodes. However, there is a caveat to this correlation: that not all stations recorded all three events (Fig. 1). SJU did not record 20 September 1995; SIO only recorded 7 January 1996; SMTC did not record 7 January 1996; and SAN did not record 17 August 1995.

Soil and basin amplitudes are generally elevated relative to rock site records, but the majority of soil and basin site amplitudes fall within 150% of the rock site amplitudes. This contrast is similar to values found by studies based on data but less than those suggested by recent simulations of ground motion in basins (Table 5).

Rogers *et al.* (1985) examines spectral ratios of data from NTS nuclear explosions recorded in the Los Angeles and San Fernando basins. Qu *et al.* (1994) and Wald and Graves (1998) compare peak amplitudes in data from the Landers mainshock to ground-motion simulations. Although the amplification factors listed for Qu *et al.* (1994) and Wald and Graves (1998) are based on only a few stations, they are consistent with the results we present here.

Spectral ratios found by Rogers *et al.* (1985) are higher than the peak amplitude ratios found by Qu *et al.* (1994), Wald and Graves (1998), and in our study. This reflects the long duration of basin-generated surface waves seen in the study by Rogers *et al.* (1985). We chose not to work with spectral ratios because many of the basin station waveforms are cut off before the coda ends. Also, we could not use a single data record as the reference for the entire southern California region, since amplitudes at rock sites can fluctuate by a factor of 2, and waveform durations vary as well, as seen in the previous section.

The simulations of Olsen *et al.* (1995) and Teng and Qu (1996) produce larger amplifications in the Los Angeles basin than are seen in the data-based studies. One possibility

Table 5

Amplification factors for Los Angeles area basins found by other researchers from data and simulations. Listings are for sources sufficiently far from the basins that surface waves have developed before hitting the basin edge. Basin abbreviations: LA, Los Angeles; SF, San Fernando; SG, San Gabriel.

Authors	Basins	<i>T</i> (sec)	Ratio	Type
Rogers <i>et al.</i> (1985)	LA, SF	>1	2–7	data spectral ratios
Qu <i>et al.</i> (1994)	LA	10	2–3 for 1D, 1–2 for 3D	data/synthetic ratio
Olsen <i>et al.</i> (1995)	LA, SF	5–10	up to 10	simulation spectral ratio
Teng and Qu (1996)	LA, SF, SG	3–10	up to 10	simulation peak amplitude
Wald and Graves (1998)	LA, SF, SG	2–17	3 in LA, 2.5 in SF	data basin/rock ratio

is source model effects. Another possibility is a combination of source and path details in the simulation. Quite possibly, though, we do not have data for the largest ground motions because of the sparseness of the network.

Wald and Graves (1998) comment that instantaneous displacement in the fault slip model used by Olsen *et al.* (1995) may increase amplitudes by a factor of 2. Teng and Qu (1996) simulate the 1857 Fort Tejon earthquake using source models with a few discrete asperities. They model only a narrow period band around 10 sec. The maximum amplitudes occur for a source model with three asperities. The timing of the asperity ruptures is not stated, but their large amplitudes may be a result of constructive interference between the rupture timing and the period of the synthetic waveforms. Qu *et al.* (1994) used a similar lumped asperity source to model the Landers mainshock and found that their results were quite sensitive to the interference created by the source model.

The largest amplifications seen in those simulation studies are highly localized, so the significance of comparisons with data from a few sites is uncertain. Our dataset may have missed points of extreme amplification in the wave field. However, the pattern of amplifications in the simulations may depend on a combination of details of the source and path models that are unique to the simulation.

At some stations, we see very clear basin-edge-generated waves. In the PAS–KIK comparison (Fig. 2a), for example, the basin waves appear to have been generated by a strong discrete pulse (visible at PAS) rather than an extended surface wave train. The discrete pulse has whiter spectral content than monochromatic surface waves and so may excite basin-edge-generated waves where a harmonic wave might not because of mismatch with the resonance frequency in the basin.

In the comparison of CRN with FUL, SAN, and OGC (Fig. 2b), a large, relatively long-period surface wave generates additional surface waves in the Los Angeles basin

with about the same dominant period. However, in other locations, there is little apparent basin response (e.g., HLN in the San Bernardino basin). Frankel (1993, 1994) has examined the response of the San Bernardino basin in 3D simulations, and data from Landers aftershocks found clear basin-generated surface waves. The dominant wave period in his records is about 2 sec. In the Ridgecrest data, the incident surface waves, as recorded at VTV and SVD, are about 10-sec period waves. The difference between the frequency excited by the basin and the dominant surface-wave frequency in Ridgecrest records may explain the lack of basin surface waves at HLN. The San Bernardino basin is a much shallower basin than the Los Angeles basin. Frankel's estimate of the deepest point is 1 km. The Los Angeles basin is up to 10 km deep (Yerkes *et al.*, 1965)).

Dispersion curves might be useful for understanding the frequency dependence of surface-wave interaction with sedimentary basins and of basin-generated surface waves. Unfortunately, basin records in the data set presented here are too often cut short or affected by long-period noise to retrieve clear dispersion relationships at longer periods (> 6 sec).

Conclusions

We are using synthetic waveforms generated with 1D models as a measuring standard to examine the variability of data throughout southern California. There is a greater variability in the amplitudes from basin site records than from rock site records. Rock, soil, and basin sites are all rather insensitive to nodes in the horizontal radiation pattern. This complicates the analysis because the nodes create singularity points in the distribution of ratios of observed and synthetic amplitudes. In particular, a Love-wave node runs through the Los Angeles basin area for all three events. When data near nodes are removed, the data at most rock sites have amplitudes within a factor of 2 of synthetic waveform amplitudes. Data from a few stations vary more (up to three times the synthetic amplitudes). Soil and basin stations vary more relative to the synthetic waveforms, with the bulk of the distribution at ratios less than 3 and a few outliers greater than 5. Soil and basin sites are also more often larger than the synthetics (higher median values).

In the Ridgecrest data set, most outliers can be explained by applying a water level of 50% to the radiation pattern. This reduces the scatter in the distributions to about the same extent as removing data within 10° of nodes. Thus, most of the outliers in the data are sites that are insensitive to the nodes, not sites that are simply larger than any other data.

Of the remaining outliers, which are not explained by insensitivity to nodes, some are over the southern Los Angeles, the Santa Barbara, and the Ventura basins. Others are south of the Los Angeles and Imperial Valley basins. These sites indicate that basin-generated surface waves are propagating through the trailing edge of these basins.

Rayleigh waves on the radial component vary in amplitude more than surface waves on other components. Our initial 1D model underpredicts the amplitude of the radial component at all types of sites. A 3-km-thick slow layer improves the fit of the synthetic waveform amplitudes, particularly at higher frequencies, but the radial component still shows the most scatter.

Acknowledgments

Reviews by S. Larsen and M. Furumura improved the original manuscript. This research was supported by the USGS Grant Number 1434-93-G-2322. This research was also supported by the Southern California Earthquake Center. SCEC is funded by NSF Cooperative Agreement EAR-8920136 and USGS Cooperative Agreements 14-08-0001-A0899 and 1434-HQ-97AG01718. SCEC Contribution Number 458. Division of Geological and Planetary Sciences Contribution Number 8496.

References

- Bard, P.-Y. and M. Bouchon (1980). The seismic response of sediment-filled valleys: Part 1. The case of incident SH waves, *Bull. Seism. Soc. Am.* **70**, 1263–1286.
- Boore, D. M., W. B. Joyner, and T. E. Fumal (1997). Equations for estimating horizontal response spectra and peak acceleration from western North American earthquakes: a summary of recent work, *Seism. Res. Lett.* **68**, 128–153.
- Devore, J. L. (1987). *Probability and Statistics for Engineering and the Sciences*, 2nd ed., Brooks/Cole, Monterey, California.
- Dibblee Jr., T. W. (1968). Geologic map of the Yucaipa quadrangle, California, U.S. Geological Survey, OFR 68-73.
- Dibblee Jr., T. W. (1989). Geologic map of the Pasadena quadrangle, Los Angeles County, California, Dibblee Geological Foundation, no. 23.
- Dibblee Jr., T. W. (1992a). Geologic map of the Oat Mountain and Canoga Park (north 1/2) quadrangles, Los Angeles County, California, Dibblee Geological Foundation, no. 36.
- Dibblee Jr., T. W. (1992b). Geologic map of the Saticoy quadrangle, Ventura County, California, Dibblee Geological Foundation, no. 42.
- Dibblee Jr., T. W. and H. E. Ehrenspeck (1993). Geologic map of the Thousand Oaks quadrangle, Ventura and Los Angeles Counties, California, Dibblee Geological Foundation, no. 49.
- Dreger, D. and D. V. Helmberger (1991). Source parameters of the Sierra Madre earthquake from regional and local body waves, *Geophys. Res. Lett.* **18**, 2015–2018.
- Frankel, A. (1993). Three-dimensional simulations of ground motions in the San Bernardino Valley, California, for hypothetical earthquakes on the San Andreas fault, *Bull. Seism. Soc. Am.* **83**, 1020–1041.
- Frankel, A. (1994). Dense array recordings in the San Bernardino Valley of Landers–Big Bear aftershocks: basin surface waves, Moho reflections, and three-dimensional simulations, *Bull. Seism. Soc. Am.* **84**, 613–624.
- Greenwood, R. B. and D. M. Morton (1991). Geologic map of the Santa Ana 1:100,000 quadrangle, California, California Division of Mines and Geology, OFR 91-17.
- Helmberger, D., R. Stead, P. Ho-Liu, and D. Dreger (1992). Broadband modeling of regional seismograms: Imperial Valley to Pasadena, *Geophys. J. Int.* **110**, 42–54.
- Ho-Liu, P. and D. V. Helmberger (1989). Modeling regional Love waves: Imperial Valley to Pasadena, *Bull. Seism. Soc. Am.* **79**, 1194–1209.
- Irvine, P. J. (1990). Landslide hazards in the Simi Valley area, Los Angeles and Ventura counties, California, California Division of Mines and Geology, OFR 90-17.
- Magistrale, H., K. McLaughlin, and S. Day (1996). A geology-based 3D

- velocity model of the Los Angeles Basin sediments, *Bull. Seism. Soc. Am.* **86**, 1161–1166.
- Morton, D. M. (1978a). Geologic map of the Fontana quadrangle, San Bernardino and Riverside counties, California, U.S. Geological Survey, OFR 78-19.
- Morton, D. M. (1978b). Geologic map of the Redlands quadrangle, San Bernardino and Riverside counties, California, U.S. Geological Survey, OFR 78-21.
- Morton, D. M. and B. F. Cox (1994). Geologic map of the Riverside East quadrangle, Riverside County, California, U.S. Geological Survey, OFR 88-754.
- Olsen, K. B., R. J. Archuleta, and J. R. Matarese (1995). Three-dimensional simulation of a magnitude 7.75 earthquake on the San Andreas fault, *Science* **270**, 1628–1632.
- Qu, J., T. L. Teng, and J. Wang (1994). Modeling of short-period surface-wave propagation in southern California, *Bull. Seism. Soc. Am.* **84**, 596–612.
- Rial, J. A. (1996). The anomalous seismic response of the ground at the Tarzana hill site during the Northridge 1994 southern California earthquake—a resonant, sliding block, *Bull. Seism. Soc. Am.* **86**, 1714–1723.
- Rogers, A. M., J. C. Tinsley, and R. D. Borchardt (1985). Predicting relative ground response, in *Evaluating Earthquake Hazards in the Los Angeles Region—An Earth-Science Perspective*, U.S. Geol. Surv. Profess. Pap. 1360, 221–247.
- Saikia, C. K. (1994). Modified frequency-wavenumber algorithm for regional seismograms using Filon-quadrature method—modeling of Lg waves in eastern North America, *Geophys. J. Int.* **118**, 142–158.
- Spudich, P., M. Hellweg, and W. H. K. Lee (1996). Directional topographic site response at Tarzana observed in aftershocks of the 1994 Northridge, California, earthquake—implications for mainshock motions, *Bull. Seism. Soc. Am.* **86**, S193–S208.
- Teng, T. L. and J. Qu (1996). Long-period ground motions and dynamic strain field of Los Angeles basin during large earthquakes, *Bull. Seism. Soc. Am.* **86**, 1417–1433.
- Wald, D. J. and R. W. Graves (1998). The seismic response of the Los Angeles basin, California, *Bull. Seism. Soc. Am.* **88**, 337–356.
- Yerkes, R. F., T. H. McCulloh, J. E. Schoellhamer, and J. G. Vedder (1965). Geology of the Los Angeles basin, California: an introduction, *U.S. Geol. Surv. Profess. Pap.* 420-A.
- Zhao, L. and D. V. Helmberger (1994). Source estimation from broadband regional seismograms, *Bull. Seism. Soc. Am.* **84**, 91–104.
- Zhu, L. and D. V. Helmberger (1996). Advancement in source estimation techniques using broadband regional seismograms, *Bull. Seism. Soc. Am.* **86**, 1634–1641.

Seismological Laboratory
California Institute of Technology
Pasadena, California 91125

Manuscript received 19 December 1997.